Inducción y Analogía en el Razonamiento Revisable

Claudio Delrieux*

Departamento de Ing. Eléctrica y Computadoras - Universidad Nacional del Sur - claudio@acm.org

PALABRAS CLAVE: Inteligencia Artificial — Razonamiento revisable — Inferencia ampliativa — Inducción

1. Introducción

La inducción puede definirse como la inferencia de una ley general a partir de casos particulares. La inducción fue de especial importancia dentro de la Epistemología y Teoría de la Ciencia, como modelo del razonamiento científico. En este proyecto buscamos incorporar inducción dentro del razonamiento revisable [8, 7, 9] La idea natural es que la ley inducida asuma la forma de un condicional derrotable más, dentro del conocimiento tentativo disponible. De esa manera, al sistema le es posible observar el conocimiento particular disponible, obtener alguna generalización al respecto, y luego generar nuevas conclusiones tentativas, compararlas, y elegir las que resulten ser más firmes. La analogía es otro mecanismo racional de inferencia utilizable para obtener conclusiones no alcanzables por la deducción. A partir de la evidencia disponible el sistema considera la posibilidad de inferir alguna propiedad que sea *proyectable* para algún término. La analogía también posee algunos problemas de fundamentación, y en consecuencia, para incluir inferencias por analogía debemos realizar algunas elecciones al diseñar nuestro sistema de razonamiento.

2. La inducción

La inducción, como mecanismo de inferencia, fue ampliamente discutido y considerado en la teoría de la ciencia, surgido de la necesidad de un método científico basado en la experimentación, la objetividad, y centrado en el *descubrimiento de leyes*. Este último punto es de especial importancia en la Inteligencia Artificial, especialmente en el área de representación de conocimiento y razonamiento (KR&R), puesto que el *conocimiento*, fruto de la investigación científica, se construye precisamente a partir la abstracción de un conjunto de observaciones acerca de un aspecto de la realidad, en lo que usualmente se denomina *teoría* (científica). Es decir, una teoría podría conceptualizarse como un conjunto de leyes que sistematiza el conocimiento de un determinado dominio. Esta manera de presentar el objeto de las ciencias experimentales nos enfrenta rápidamente a los problemas más profundos e interesantes de la Epistemología: cómo se construye el conocimiento científico, cómo se justifica dicho conocimiento, cómo se contrastan sus virtudes, y cómo y por qué se modifica por otro conocimiento más satisfactorio. Los modelos de razonamiento usuales en KR&R parecen restringidos a la segunda y tercer tarea (y con ciertas limitaciones), mientras que la primera y la cuarta no han sido objeto de estudio pese a su evidente importancia. De ello nos ocuparemos en este trabajo.

*Parcialmente financiado por la SECyT-UNS

Ninguna justificación para la inducción, al igual que para cualquier otra inferencia ampliativa, puede brindar a este esquema una solidez comparable a la deducción. Una de las primeras y más importantes defensas de la inducción consistió en afirmar que se trata de un patrón de inferencia *probabilístico*. Este tipo de caracterizaciones reaparece cada tanto en la Lógica y en la Inteligencia Artificial, con algunas u otras modificaciones, en la forma de lógicas inductivas [6], Bayesianismo [5], razonamiento evidencial [4] etc. Además, en Inteligencia Artificial, muchos métodos de aprendizaje automático, algunas caracterizaciones del *commonsense reasoning* y otras representaciones computacionales del descubrimiento científico se basan en forma lisa y llana en utilizar a la inducción como proceso generativo de hipótesis, en un marco general de prueba y error. En estas condiciones, la suposición es que una inferencia inductiva transfiere parcialmente la verdad de las premisas a la regla inferida, si el contexto se mantiene.

3. Inducción en el Razonamiento Revisable

Para incorporar la inducción de reglas derrotables dentro de nuestro enfoque, consideramos a la inducción como la generación de clases de equivalencia parciales. Por lo tanto, nuestro modelo será el de inducir cautamente una regla derrotable $a(X) \succ\!\!- b(X)$ a partir de una suficiente constatación de casos t_i en los cuales siempre que se observaba $a(t_i)$, también se observaba $b(t_i)$.

EJEMPLO 1 *Supongamos tener como teoría* $\mathcal{T} = \{a(X) \succ\!\!- c(X), c(X) \succ\!\!- \neg b(X)\}$, *y como evidencia observada* $E = \{a(t_1), a(t_2), b(t_1), b(t_2), b(t_3), a(s)\}$. *¿Qué podemos predecir acerca de* $b(s)$*? Por un lado tenemos el argumento* $\{a(s) \succ\!\!- c(s), c(s) \succ\!\!- \neg b(s)\}$, *el cual nos permite concluir* $\neg b(s)$. *Pero por otro lado, a partir de la evidencia podemos proponer la siguiente inducción*

$$\frac{\begin{array}{c} a(t_1), a(t_2) \\ b(t_1), b(t_2), b(t_3) \end{array}}{a(X) \succ\!\!- b(X).}$$

La nueva regla inducida provee un argumento para $b(s)$*, el cual es más directo que el argumento anterior para* $\neg b(s)$*.*

Este ejemplo ilustra la utilidad de poder inferir inductivamente nuevas reglas derrotables, pero también pone a las claras algunos puntos discutibles que pueden presentarse:

- El antecedente de la inducción utiliza una parte de la evidencia. ¿Por qué no se utiliza otra parte de la evidencia para inducir otra regla? Por ejemplo

$$\frac{\begin{array}{c} b(t_1), b(t_2) \\ a(t_1), a(t_2), a(s) \end{array}}{b(X) \succ\!\!- a(X).}$$

- ¿Cuántos átomos "confirmatorios"en la evidencia son suficientes para "disparar" una inferencia inductiva?
- ¿Tiene la regla derrotable inducida la misma valoración epistémica que las demás reglas en $\mathcal{T}$? Si no fuera así ¿sigue siendo adecuado que una regla más específica pero menos fuerte derrote a un argumento más fuerte pero menos específico?
- ¿Pueden ser utilizados los literales que se generan por medio de argumentos para inducir nuevas reglas?

Es poco probable que estas preguntas tengan una respuesta única en todos los contextos posibles. Evidentemente, la búsqueda de respuestas para estas y otras cuestiones abre el camino para recorrer nuevos campos de investigación. Más cauto es asumir que cada una de ellas representa un parámetro a elegir en el proceso de *diseñar* un sistema de razonamiento. En este trabajo vamos a asumir una elección particular de parámetros para la inferencia inductiva, quedando abierta la posibilidad de estudiar comparativamente los resultados de modificar estas elecciones.

Definición 1 *Sea E un conjunto de evidencia compuesto por literales de base. Un* caso confirmatorio *para $a(X) \succ\!\!\!\!- b(X)$ en E es un término t tal que $a(t)$ y $b(t)$ están en E.*

Un caso refutatorio *para $a(X) \succ\!\!\!\!- b(X)$ en E es un término t tal que $a(t)$ está en E, pero $\neg b(t)$ está en E.*

Un caso desconocido *para $a(X) \succ\!\!\!\!- b(X)$ en E es un término t tal que $a(t)$ está en E, pero $b(t)$ no está en E.*

Definición 2 *Dados dos predicados a, b en E, sea T_a el conjunto de términos tales que $a(t) \in E$ si $t \in T_a$. Del mismo modo, sea T_b el conjunto de términos tales que $b(t) \in E$ si $t \in T_b$.*

Llamemos ahora

$$\begin{array}{rcl} A & = & T_a/T_b, \\ B & = & T_b/T_a, \\ C & = & T_a \cap T_b. \end{array}$$

Los términos en C son casos confirmatorios para $a(X) \succ\!\!\!\!- b(X)$ y para $b(X) \succ\!\!\!\!- a(X)$. Los términos en A son casos desconocidos para $a(X) \succ\!\!\!\!- b(X)$, mientras que los términos en B son casos desconocidos para $b(X) \succ\!\!\!\!- a(X)$. Por lo tanto,

si $|C| > |B|$ y no hay ningún caso refutatorio, se infiere $a(X) \succ\!\!\!\!- b(X)$,
si $|C| > |A|$ y no hay ningún caso refutatorio, se infiere $b(X) \succ\!\!\!\!- a(X)$.

Ejemplo 2 *(Ejemplo 1 revisitado). Tenemos los siguientes conjuntos:*
$T_a = \{t_1, t_2, s\}$,
$T_b = \{t_1, t_2, t_3\}$,
$A = \{s\}$, $B = \{t_3\}$, $C = \{t_1, t_2\}$.

En este caso se confirma $a(X) \succ\!\!\!\!- b(X)$ con los términos t_1 y t_2, mientras que el término s es un caso desconocido. Por lo tanto la regla $a(X) \succ\!\!\!\!- b(X)$ es inferida por inducción.

También es posible observar que la regla $b(X) \succ\!\!\!\!- a(X)$ se confirma con los términos t_1 y t_2, mientras el término t_3 es un caso desconocido. Por lo tanto la regla $b(X) \succ\!\!\!\!- a(X)$ es también inferida por inducción.

Entonces tenemos ahora

$$\mathcal{T}' = \{a(X) \succ\!\!\!\!- c(X), c(X) \succ\!\!\!\!- \neg b(X) a(X) \succ\!\!\!\!- b(X), b(X) \succ\!\!\!\!- a(X)\}$$

y como evidencia observada

$E = \{a(t_1), a(t_2), b(t_1), b(t_2), b(t_3), a(s)\}$.

La regla $a(X) \succ\!\!\!\!- b(X)$ provee un argumento no derrotado para concluir $b(s)$ a partir de $a(s)$, y la regla $b(X) \succ\!\!\!\!- a(X)$ provee un argumento no derrotado para concluir $a(t_3)$ a partir de $b(t_3)$.

La definición anterior no especifica qué hacer en el caso de que $|C| > |A| > |B|$ o $|C| > |B| > |A|$. Tal vez una política inductiva adicional en cualquiera de esos casos sea inferir solamente la regla que contempla la menor cantidad de casos desconocidos.

4. Razonamiento por analogía

Si la inducción puede pensarse como una inclusión o clase de equivalencia parcial, entonces la analogía puede verse como un isomorfismo parcial entre dos relaciones:

$$\begin{array}{l}
a_1(s), a_2(s), \cdots, a_n(s), a_{n+1}(s), \cdots, a_{n+k}(s) \\
b_1(t), b_2(t), \cdots, b_n(t) \\
a_1(X) \underset{\text{iso}}{\longleftrightarrow} b_1(X) \\
a_2(X) \underset{\text{iso}}{\longleftrightarrow} b_2(X) \\
\vdots \\
a_n(X) \underset{\text{iso}}{\longleftrightarrow} b_n(X) \\
\hline
b_{n+1}(t), \cdots, b_{n+k}(t).
\end{array}$$

Si hay dos (tuplas de) términos, s (source o fuente) y t (target o destino), para los que se conoce que comparten respectivamente las propiedades isomorfas o proyectables a_i y b_i ($i \in \{1..n\}$), entonces, si s posee propiedades adicionales a_j ($j \in \{n+1..n+k\}$), entonces se infiere por analogía que t posee las propiedades isomorfas o proyectables adicionales b_j. Al igual que en la inducción, cabe plantear una serie de cuestiones que permitan discriminar una buena analogía de otra no tan buena.

- ¿Cuántos individuos s o casos "confirmatorios" en la evidencia son suficientes para "disparar" una inferencia por analogía?
- ¿Cuántas propiedades a_i deben compartir los individuos s y t para poder inferír que t posee propiedades que s posee?
- Si aceptamos que t posee propiedades que s posee ¿Cuál es la importancia epistémica de dicha conclusión?
- ¿Cuándo y por qué se decide que una propiedad es proyectable?
- ¿Pueden ser utilizados los literales que se generan por medio de argumentos para realizar analogías?

Al igual que en la Sección anterior con la inducción, realizaremos aquí una serie de elecciones posibles, dejando abierto el campo para estudiar las diferentes posibilidades en trabajos futuros. Parece razonable exigir que la cantidad de casos confirmatorios sea mayor que la de los casos desconocidos, y que no haya ningún caso refutatorio. Al mismo tiempo, parece difícil decidir por qué razón la cantidad de propiedades compartidas por los individuos s y t debe ser un número determinado. También dejamos el status epistemológico de los literales inferidos por analogía como si fuesen equivalentes a la evidencia misma (algo que podría modificarse utilizando plausibilidad [1]). Podríamos resumir suscintamente la manera de operar de nuestro sistema de razonamiento.

1. Nuestro sistema parte de un conjunto de evidencia E que consta de literales de base, una teoría $\mathcal{T}$ que consta de un conjunto de reglas derrotables, y un contexto de conocimiento firme $\mathcal{K}$.
2. En un primer paso, es posible generar nuevas reglas derrotables por inducción en E, y también generar nuevos literales por analogía en E. De esa forma, tenemos un nuevo conjunto de evidencia ampliada E', y una teoría ampliada $\mathcal{T}'$.

3. Con E' y $\mathcal{T}'$ es posible, entonces, construir argumentos y generar nuevas conclusiones.

Un problema interesante surge de la interacción entre los pasos 2 y 3. En particular pueden encontrarse casos sencillos donde el conocimiento obtenido por inducción produce conclusiones contradictorias con el conocimiento obtenido por analogía.

5. Conclusiones y trabajo futuro

Presentamos dos reglas de inferencia que permiten incluir, respectivamente, induccion y analogía dentro del razonamiento revisable. En ambos casos se discutieron las condiciones dentro de las cuales parece razonable aplicar este tipo de razonamiento, junto con ejemplos y algunas de sus limitaciones. En ambos casos se pudo mostrar en qué medida el proceso de incluir estas reglas, así como las condiciones particulares que permiten aplicarlas, es en definitiva un proceso de diseño de un sistema de razonamiento con las características particulares adecuadas en un determinado contexto. Actualmente estamos trabajando en la búsqueda de criterios de aceptación que vayan más allá de la mera búsqueda de casos confirmatorios y refutatorios. Otra linea de investigación consiste en estudiar los criterios de inter-derrota entre argumentos generados utilizando diferentes patrones ampliativos, no solo inducción y analogía, sino también abducción (ver [3]). Por último, es importante también considerar la plausibilidad de la base de información en cada uno de estos mecanismos, lo cual permitiría tener un criterio adicional de inter-derrota (ver [1]).

Referencias

[1] Claudio Delrieux. Nonmonotonic Reasoning Under Uncertain Evidence. In Fausto Giunchiglia, editor, *Artificial Intelligence: Methodology, Systems and Applications*, pages 195–204. Springer, ISSN 3-540-64993, Lecture Notes in Artificial Intelligence 1480, 1998.

[2] Claudio Delrieux. The Rôle of Defeasible Reasoning in the Modelling of Scientific Research Programmes. In *Proceedings of the IC-AI 2001 Conference*, pages 861–868, CSREA Press, ISBN 1-892512-81-5, 2001.

[3] Claudio Delrieux. Abductive Inference in Defeasible Reasoning: a Model for Research Programmes. *Journal of Applied Logic*, 2:409–437, 2004.

[4] Judea Pearl. Evidential Reasoning using Stochastic Simulation. *Artificial Intelligence*, 32(3):245–257, 1987.

[5] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, California, 1988.

[6] John Pollock. How to Reason Defeasibily. *Artificial Intelligence*, 57(1):1–42, 1992.

[7] H. Prakken and G. Sartor. Argument-Based Logic Programming with Defeasible Priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.

[8] Guillermo R. Simari and Ronald P. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53(2-3):125–158, 1992.

[9] Bart Verheij. Argue!, an Implemented System for Computer-Mediated Defeasible Argumentation. In *Proceedings of the Tentht Netherlands/Belgium Conference on Artificial Intelligence*, pages 57–66, Netherlands, 1998. NAIC'98, Amsterdam.